

Washington University School of Medicine Digital Commons@Becker

Open Access Publications

2007

Fine mapping of mouse QTLs for fatness using SNP data

Armin O. Schmitt
Humboldt-University Berlin

Hadi Al-Hasani
German Institute for Human Nutrition

James M. Cheverud
Washington University School of Medicine

Daniel Pomp
University of Nebraska - Lincoln

Lutz Bünger
Scottish Agricultural College

See next page for additional authors

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation

Schmitt, Armin O.; Al-Hasani, Hadi; Cheverud, James M.; Pomp, Daniel; Bünger, Lutz; and Brockmann, Gudrun A., "Fine mapping of mouse QTLs for fatness using SNP data." *OMICS: A Journal of Integrative Biology*.11,4. 341-350. (2007).
http://digitalcommons.wustl.edu/open_access_pubs/4745

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.

Authors

Armin O. Schmitt, Hadi Al-Hasani, James M. Cheverud, Daniel Pomp, Lutz Bünger, and Gudrun A. Brockmann

Fine Mapping of Mouse QTLs for Fatness Using SNP Data

ARMIN O. SCHMITT,¹ HADI AL-HASANI,² JAMES M. CHEVERUD,³
DANIEL POMP,⁴ LUTZ BÜNGER,⁵ and GUDRUN A. BROCKMANN¹

ABSTRACT

Quantitative trait loci (QTLs), as determined in crossbred studies, are a valuable resource to identify genes responsible for the corresponding phenotypic variances. Due to their broad chromosomal extension of some dozens of megabases, further steps are necessary to bring the number of candidate genes that underlie the detected effects to a reasonable order of magnitude. We use a set of 13,370 SNPs to identify informative haplotype blocks in 22 mouse QTLs for fatness. About half of the genes in a typical QTL overlap with haplotype blocks, which are different for the two base mouse lines, and which, thus, qualify for further analysis. For these genes we collect four more pieces of evidence for association with fat accumulation, namely (1) homology to genes identified in a *Caenorhabditis elegans* knock-out experiment as fat decreasing or fat increasing, (2) the overexpression of the genes in mouse fat, liver, muscle, or hypothalamus tissues, (3) the occurrence of a gene in several independently found QTLs, and (4) the information provided by gene ontology, to achieve a ranked list of 131 candidate genes. Ten genes fulfill three or four of the above sketched criteria and are discussed briefly, 121 further genes fulfilling two criteria are provided as on-line material. Viewing the genomic region of fatness-related QTLs under several different aspects is appropriate to assess the many thousands of genes that reside in such QTLs and to produce lists of more robust candidate genes.

INTRODUCTION

QUANTITATIVE TRAIT LOCI (QTLs) often represent the entry point into a whole series of successive analyses to pinpoint genes influencing a phenotypic trait of interest. Obesity and body weight are among the most analysed traits in humans and model animals. Although many knock-out mouse models and cell tissue experiments have shown that various genes regulate adipocyte differentiation and proliferation (Sakai et al., 2007; Yamauchi et al., 2007), energy partitioning (Ren et al., 2007) glucose homeostasis, and carbohydrate and lipid metabolism (Gao et al., 2007; Le Bacquer et al., 2007), only few genes have shown as-

¹Institute for Animal Sciences, Humboldt-Universität zu Berlin, Berlin, Germany.

²German Institute for Human Nutrition, Bergholz-Rehbrücke, Germany.

³Department of Anatomy & Neurobiology, Washington University School of Medicine, St. Louis, Missouri.

⁴Animal Science Department, University of Nebraska, Lincoln, Nebraska.

⁵Sustainable Livestock Systems, Scottish Agricultural College, Penicuik, Midlothian, UK.

sociation to naturally occurring obesity in model animal and human population studies (Brockmann and Bevoja, 2002). The major reason for this is the complex nature of body weight regulation in most cases of obesity. Many genes have relatively small effects and interaction among each other, and also responses to environmental factors contribute to the phenotype. Thus, it remains a difficult task to identify the genes that are responsible for obesity epidemic in humans.

So far, more than 30 independent QTL studies have been performed to detect almost 100 QTLs associated with fatness in mice (Wuschke et al., 2007) or humans (Perusse et al., 2005). Unfortunately, QTLs typically span large genomic regions: QTLs extending over half a chromosome are not unusual. About two-thirds of the mouse genome are covered by at least one fatness-associated QTL. Thus, the need to fine map QTLs to identify the underlying gene or genes is obvious, as pointed out clearly in a recent review (Flint et al., 2005). This review also puts forward a comprehensive list of strategies how this can be accomplished. Widely used fine mapping approaches include the production of congenic lines (Jerez-Timaure et al., 2005; Klein, 1978; Stylianou et al., 2004), recombinant inbred lines (Williams et al., 2001), and advanced intercross lines (Darvasi and Soller, 1995). The combination of genetic dissection of a QTL by recombination events with the analysis of individual gene expression profiles was shown to permit the mapping of differentially expressed positional candidate genes (Brown et al., 2005; Hitzemann et al., 2003; Mehrabian et al., 2005). Correlations between phenotypic traits and genomic *loci* were established by combining QTL mapping and expression analysis (Lan et al., 2006; Stylianou et al., 2005). The general utility and applicability to use SNP data to narrow down regions of candidate genes was previously demonstrated by means of an SNP data set of similar size as the one we use in our study (Pletcher et al., 2004). In other studies, QTL-specific genotype analysis was used to identify *Apoa2* as the gene controlling high-density lipoprotein levels (Wang et al., 2004) or *Insig2* as a susceptibility gene for plasma cholesterol level (Cervino et al., 2005). If the haplotypes of the base lines of a crossbred population are known within the QTLs, haplotype regions can be identified that differ between the lines. These regions might harbor the gene or genes underlying the QTL effect, and thus the observed phenotypic variance. The rationale of this approach was sketched in a recent article (DiPetrillo et al., 2005). In this work, we first identify all genes that lie in those parts of QTLs where genotypes differ. Then we subject those genes to further bioinformatics analyses to gather more evidence for the involvement of the genes in fat accumulation. We show that by a combination of several filters, including homology searches against known fat-related genes, gene expression, and Gene Ontology annotation, the number of genes can be efficiently reduced.

MATERIALS AND METHODS

QTL and genotyping data

We first identified eight QTL studies in the literature with mouse crosses where both divergent strains were genotyped. Altogether, 22 QTLs for fat accumulation or fat percentage were identified in these eight studies. Then we extracted the chromosomal locations of these 22 QTLs from a metastudy in which all published mouse QTLs are presented and made comparable (Wuschke et al., 2007). We used the Wellcome-Complex Trait Consortium (CTC) Mouse Strain SNP Genotype Set Build 34 for our analysis (<<http://www.well.ox.ac.uk/mouse/INBREDS>>). This genome-wide data set includes 480 mouse lines that were genotyped at 13,370 SNP positions. The overwhelming part of the SNPs are spaced around 230 kbp, which is the average distance, apart. In very few cases, however, distances of more than one megabase occur. In order to find QTL specific informative haplotypes (i.e., those parts of a QTL where genotypes differ) we first filtered out SNPs within QTLs and tagged all informative SNPs, that is, SNPs with different genotype between the lines. Then we searched for continuous stretches of informative SNPs, which we termed informative haplotype blocks (IHTBs). In order to find longer stretches we tolerated isolated non-informative SNPs (i.e., SNPs with neighboring informative SNPs) in a stretch of informative SNPs.

Homologues to C. elegans genes

In a genome wide RNAi analysis of *Caenorhabditis elegans* 413 fat-decreasing and 133 fat-increasing genes were identified (Ashrafi et al., 2003). Mouse genes were considered homologous to one such gene

if they produced a hit in a Blast search (Altschul et al., 1997) (blastall version 2.2.8; DNA vs. DNA in the six-frame translation option tblastx) with an E-value of 0.00001 or smaller and a sequence identity of 30% or more on the protein level. In the blastall mode tblastx both query and target DNA sequence are translated into all six possible reading frames and are then aligned. This way, also distant relationships between sequences can be detected.

Gene expression analysis

The mRNA expression level of candidate genes was calculated for four tissues (liver, hypothalamus, muscle, and adipocyte) from mice in normal physiological state based upon an expression dataset provided by the Genomics Institute of the Novartis Research Foundation (Su et al., 2002). The expression strength was determined from the raw intensity data by means of the *expres* function from the R-package *affy* (Gautier et al., 2004; R Development Core Team, 2005). A gene is considered strongly expressed in a tissue if its expression level lies above the 90th percentile on both chips that were hybridized individually to a tissue.

Gene ontology annotation

We considered a gene as known to be involved in fat accumulation or metabolism if its GO description (Ashburner et al., 2000) as provided by BioMart (Kasprzyk et al., 2004) contained one or more of the following terms: *metabol**, *mitochondr**, *lipid*, *fat*, *carboxy**, *nutrient*, *dopamin**, *peroxisom**, where the asterisk stands for any other string of text. All mouse genes were downloaded into a file with their GO descriptions via the BioMart homepage (<http://www.ensembl.org/Multi/martview>) and the genes characterized by above-mentioned keywords were then searched in that file using the Unix utility *grep* function.

Bioinformatics work

All bioinformatics work was done using the R-statistics package (R Development Core Team, 2005) and *bioperl-modules* (Stajich et al., 2002). The Ensembl gene identifiers served as reference in this study.

RESULTS

We could match 22 QTLs for fatness from eight different studies with the corresponding SNP genotype data of the mouse lines used in these QTL studies (see Table 1). These 22 QTLs cover 573 out of the 2600 megabases of the murine genome (22%); 2.8% of the genome is covered by two, and about one megabase is covered by three different QTLs (Fig. 1). As the SNP density across the entire genome is reasonably high (on average, one SNP every 230 kb), it is very unlikely that the gene or the genes responsible for the susceptibility toward fat accumulation resides in a region of the QTL that does not differ between the two lines. Thus, further analysis can be limited to genes located in QTL regions of different haplotypes between the two lines. We dissected each QTL into intervals where the two lines that contributed to the mapping of the QTL differed in haplotype. Such a QTL-specific haplotype analysis for all 22 QTLs yields 225 informative haplotype blocks. These 225 IHTBs comprise between 73 kb and 11.4 megabases in length, with a median of 0.9 megabases. The 22 QTLs were dissected into on average eight IHTBs (minimum 2, maximum 27). The rationale of this procedure is that the causative gene has to lie in a QTL region of divergent haplotypes. QTLs are on average reduced by about 50% in size by such a filtering. Likewise, the number of genes residing in such informative parts of QTLs is also roughly halved (5822 unique genes in QTLs vs. 3106 unique genes in informative parts of QTLs). We identified 73 mouse genes homologous to fat-increasing worm genes and 224 genes homologous to fat-decreasing worm genes. We further found 99 genes highly expressed in hypothalamus, 80 genes in muscle, 91 in adipocyte tissue, and 64 in liver. Two hundred sixty-eight genes had a gene ontology description that indicated involvement in metabolism and 287 genes were found to be located in two or more different QTLs. We consider the fulfilment of any of the above four criteria, homology, expression, Gene Ontology annotation, and multiple QTL coverage, as a piece of evidence. We therefore sum up the pieces of evidence to yield a score, and present all genes with a score of three or more in Table 2.

TABLE 1. MOUSE CROSSES USED FOR THE QTL-SPECIFIC HAPLOTYPE ANALYSIS

| Cross and reference | Name of | | | Position of QTL | | | #genes in | | | Reduction of #genes in % |
|---|---------|--------------|-------|-----------------|-------|--------|-----------|-----|------|-----------------------------|
| | Chrom | QTL | Trait | Start | End | Length | #IHTB | QTL | IHTB | |
| C57BL/6J \times 129S1/SvImJ (Ishimori, 2004) | 8 | <i>Obq16</i> | F% | 85.8 | 108.4 | 22.6 | 12 | 287 | 99 | 66 |
| SM/J \times NZO/HILt (Taylor, 2001) | 1 | <i>Obq7</i> | F% | 42.0 | 70.7 | 28.7 | 6 | 229 | 74 | 68 |
| | 1 | <i>Obq8</i> | F% | 121.9 | 162.3 | 40.4 | 13 | 317 | 167 | 47 |
| | 1 | <i>Obq9</i> | F% | 156.6 | 171.6 | 15.0 | 4 | 156 | 41 | 74 |
| | 2 | <i>Obq10</i> | F% | 93.5 | 121.0 | 27.5 | 17 | 273 | 176 | 36 |
| | 5 | <i>Obq11</i> | F% | 4.25 | 28.4 | 24.15 | 4 | 189 | 57 | 70 |
| | 5 | <i>Obq12</i> | F% | 38.1 | 56.4 | 18.3 | 7 | 100 | 26 | 74 |
| | 6 | <i>Obq13</i> | F% | 48.9 | 56.8 | 7.9 | 2 | 87 | 60 | 31 |
| | 6 | <i>Obq14</i> | F% | 93.7 | 106.8 | 13.1 | 3 | 49 | 18 | 63 |
| | 7 | <i>Obq15</i> | F% | 74.5 | 98.8 | 24.3 | 11 | 228 | 145 | 36 |
| | 17 | <i>Obq4b</i> | F% | 10.4 | 26.2 | 15.8 | 4 | 318 | 98 | 69 |
| NZOHILt \times NZOLt (Reifsnyder, 2000) | 12 | <i>Nzoq2</i> | F% | 87.3 | 111.7 | 24.4 | 12 | 278 | 159 | 43 |
| AKR/J \times SWR/J (West, 1994) | 4 | <i>Do1</i> | F% | 97.7 | 126.5 | 28.8 | 15 | 401 | 186 | 54 |
| AKR/J \times C57L/J (Taylor, 1997) | 2 | <i>Obq3</i> | F% | 60.6 | 128.0 | 67.4 | 27 | 891 | 750 | 16 |
| NZB/BINJ \times SM/J (Lembertas, 1997) | 2 | <i>Bfq1</i> | fw | 136.4 | 151.8 | 15.4 | 4 | 134 | 84 | 37 |
| F \times L (Horvat, 2000) | 2 | <i>Fob1</i> | F% | 54.4 | 94.6 | 40.1 | 9 | 572 | 164 | 71 |
| | 12 | <i>Fob2</i> | F% | 7.8 | 33.5 | 25.7 | 7 | 181 | 157 | 13 |
| | 15 | <i>Fob3</i> | F% | 9.4 | 77.5 | 68.1 | 22 | 452 | 263 | 42 |
| | X | <i>Fob4</i> | F% | 73.0 | 113.2 | 40.3 | 3 | 265 | 95 | 64 |
| C57BL/6J \times KK/HILt (Taylor, 1999) | X | <i>Obq6</i> | F% | 44.0 | 62.9 | 18.9 | 6 | 144 | 75 | 48 |
| | 7 | <i>KK</i> | F% | 40.9 | 81.1 | 40.2 | 13 | 621 | 214 | 66 |
| | 9 | <i>Obq5</i> | fw | 14.8 | 55.0 | 40.2 | 23 | 580 | 295 | 49 |

All eight mouse crosses whose base lines were genotyped are listed. Their QTLs are reproduced along with their chromosomal locations and the number of genes found in them. Informative haplotype blocks (IHTB) are those parts of QTLs where the lines' genotypes differ. IHTBs cover on average half of a QTL, and, thus, the number of genes in IHTBs is also half the number in the whole QTL. Start, end, and length of QTLs are given in megabases. Abbreviations: F%, fat percentage; fw, fat weight; IHTB, informative haplotype block.

DISCUSSION

We presented an approach how genotyping data can be used to reduce the large number of genes in a typical QTL for fatness, and how further information can help to rank candidate genes. We found that the 22 QTLs of eight mouse crosses under consideration are almost randomly distributed across the murine genome. It is interesting to note that the genome is maximally covered by QTLs from only three out of eight different crosses. There is no chromosomal region that is covered by QTLs from the majority of the eight crosses under consideration. This weak congruence between QTLs from different crosses shows that fat accumulation (and its pathological variants adiposity and obesity) is, despite its phenotypic similarity, a very multifaceted trait and can be caused by a manifold of genes. In the polygenic case, the sets of alleles causing fatness can be very different between fat lines (Wagener et al., 2006).

Our approach is based upon the exclusion principle: a chromosomal region that is identical for two divergent phenotypes cannot harbor genes responsible for the divergence in phenotype. The reduced number of genes resulting from such a filtering should therefore be enriched with genes that might cause abnormal fat accumulation. Correspondingly, genes that are not involved in the process of fat accumulation should

FINE MAPPING OF MOUSE QTLs

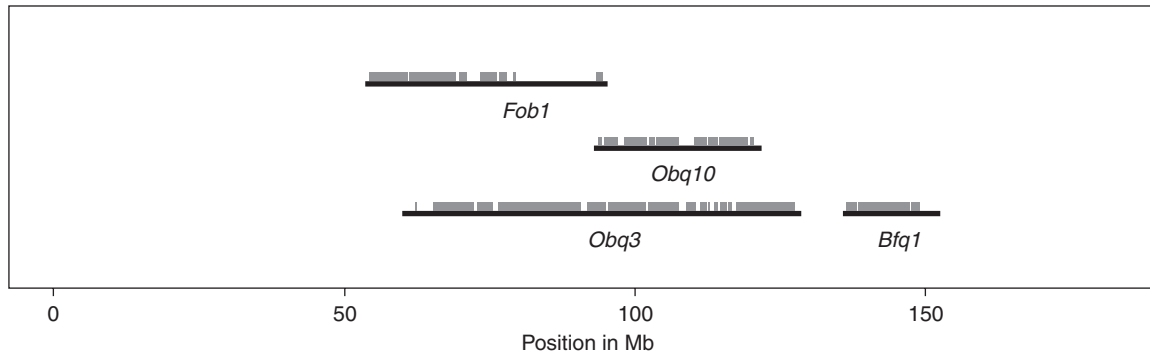


FIG. 1. Four genotyped QTLs on the murine chromosome 2. Shown are the four fatness related murine QTLs that were found on chromosome 2 (black lines). Only QTLs from studies where both lines were genotyped by the Complex Trait Consortium are presented. The QTLs overlap to a large extent; along one megabase there is even threefold overlap. The little gray boxes indicate those parts of the QTLs where genotypes differ between the two mouse lines.

be considerably reduced in number by such an analysis. We showed that, in such a way, a reduction of the number of possible candidate genes by about 50% can be achieved. Of course, the presented strategy has its limitations. The dataset cannot represent line-specific new spontaneous mutations that might also be of interest for the process of fat deposition. Furthermore, very small haplotype blocks, that is, shorter than the distance between two SNP markers will not be detected.

Despite this drastic reduction, the information about the haplotype block structure in the QTL regions leaves the number of genes still in the order of hundreds for a typical QTL study. Thus, it is obvious that additional criteria have to be applied to reduce this number further. In this work we drew upon the idea that the association of a gene with obesity should be further supported by more features. We therefore included into our analysis homology to fat-decreasing or -increasing worm genes, high expression in four tissues, and the annotation provided by the gene ontology database. Whereas fulfilment of the used criteria is neither sufficient nor necessary for a gene to cause fat accumulation, it is sound to believe that its likelihood for involvement increases with the number of points of evidence. We generated a list of genes that meet at least three (out of four) criteria. This list comprises both genes that were already described as fatness related (e.g., *Ganc*) but also barely characterized genes (e.g., the Riken clone C330002I19Rik). Among this list of 10 genes, Calpain3 (*Capn3*) is the only one to meet all four criteria. *Capn3* is a calcium-dependent neutral cysteine proteinase with several functions including signal transduction, calcium binding, and hydrolase activity. The expression of the *Capn3* gene in skeletal muscle is associated with body fat content and measures of insulin resistance (Walder et al., 2002). The appearance of the well-described gene *Ganc* as top scoring in our list of candidate genes makes us confident that the approach that we chose is justified and that the other genes that we present, for example, *Atp5g3*, *Ivd*, or *Plcb2*, can also be considered as candidate genes. We mention in passing that *Ganc* and *Capn3* share the same Ensembl gene identifier, ENSMUSG00000062646. Because the Ensembl gene identifier served as a reference in this study, they were presented as identical genes in Table 2.

Among the 10 high-scoring genes there are two members of the solute carrier family 25, *Slc25a29*, and *Slc25a12*. Solute carrier family 6 members 3 and 14 were identified as obesity related in previous studies (Durand et al., 2004; Epstein et al., 2002; Suviolahti et al., 2003). *Slc25a12*, which is also known as *Aralar1*, was found to determine glucose metabolic fate, mitochondrial activity, and insulin secretion in beta cells (Rubi et al., 2004). Another mitochondrial gene is *Atp5g3*, an ATP synthase with lipid-binding properties. It is primarily expressed in olfactory sensory neurons (Yu et al., 2005). The isovaleryl coenzyme A dehydrogenase gene, *Ivd*, plays a role in electron transport and metabolic processes. *Ivd* was found to be upregulated in the heart tissue of rats upon starvation, but downregulated upon a high-fat diet (Nagao et al., 1993). The phospholipase C beta 2 (*Plcb2*) contributes to the intracellular signaling cascade and lipid catabolic processes. The acyl-CoA synthetase bubblegum family, *Acsbg1*, is a gonadotropin-regulated long-

TABLE 2. TOP-RANKING CANDIDATE GENES

| Gene | Description | GO annotation | Homology ^a | Expression in ^b | Overlapping QTLs |
|----------------------|---|-------------------------------------|-----------------------|----------------------------|----------------------------|
| <i>Capn3</i> (Ganc) | calpain 3 | carbohydrate metabolism | up | M | <i>Obq1</i> , <i>Obq3</i> |
| <i>Atp5g3</i> | ATP synthase, mitochondrial complex | lipid binding | — | HLAM | <i>Obq3</i> , <i>Fob1</i> |
| <i>Slc25a29</i> | solute carrier family 25, member 29 | mitochondrion | down | M | <i>Nzoq2</i> |
| <i>Slc25a12</i> | solute carrier family 25, member 12 | mitochondrial inner membrane | down | — | <i>Obq3</i> , <i>Fob1</i> |
| <i>Ivd</i> | isovaleryl coenzyme A dehydrogenase | isovaleryl-CoA dehydrogenase | down | — | <i>Obq10</i> , <i>Obq3</i> |
| <i>Oshpl9</i> | oxysterol binding protein-like 9 | lipid transport, steroid metabolism | down | — | <i>Dol</i> |
| <i>Acsbg1</i> | acyl-CoA synthetase bubblegum family | long-chain fatty acid metabolism | down | H | <i>Obq5</i> |
| <i>C330002119Rik</i> | MKIAA1250 protein (Fragment) | fixation of carbon dioxide | down | H | <i>Fob2</i> |
| <i>Plcb2</i> | phospholipase C, beta 2 | lipid metabolism | — | HLAM | <i>Obq10</i> , <i>Obq3</i> |
| <i>Gchlfr</i> | GTP cyclohydrolase I feedback regulator | neurotransmitter metabolism | — | L | <i>Obq10</i> , <i>Obq3</i> |

Those genes are listed that have at least three scoring points out of four (GO-annotation, homology to fat increasing or decreasing *C. elegans* genes, expression, and presence in multiple QTLs; for details, see the Methods section). All genes lie in regions of fatness-related murine QTLs where genotypes differ between the base mouse lines. For a description of QTLs, see Table 1.

^aUp, homology to fat increasing *C. elegans* gene; down, homology to fat decreasing *C. elegans* gene

^bH, hypothalamus; L, liver; A, adipocyte tissue; M, muscle.

chain acyl-CoA synthetase that is essential for the synthesis of long and very long fatty acids (Pei et al., 2006; Steinberg et al., 2000). The oxysterol binding protein-like 9 protein (*Ospbl19*) is involved in the transport of lipids and the steroid metabolism. The protein, encoded by the gene with the identifier C330002I19Rik, has not been characterized yet.

The more comprehensive supplementary list of filtered candidate genes comprises genes involved in diverse gene ontology categories. Among the genes are several genes encoding proteins with transcription factor activity (*Wt1*, *Rai14*, *Zfp202*, *Nr4a2*, *Lass6*, *Pax6*, *Elf5*, *Nr2f2*, NP_899084.2, *Neurod1*, *Glis1*, *Mitf*, *Zscan2*, and *Sp5*), which might regulate cascades of genes involved in different metabolic and regulatory pathways. In experimental studies and association analyses, it will be interesting to see which of the selected candidate genes contribute to adiposity in different models.

None of the genes on the list of candidate genes lies in the small region of one megabase on chromosome 2 where three QTLs overlap. This region seems to play no role for the obese phenotypes in the corresponding crosses. But 6 out of the top 10 genes reside in the two regions on chromosome 2, which are covered by two QTLs. Beside the crosses that we could use for this analysis, additional crossbred mouse populations have shown that several genes contributing to body weight and fatness are located on chromosome 2. At least three different QTLs can be expected on this chromosome. QTLs in the crosses M16i \times L6, which were extracted in congenic lines from a cross between M16i and C57BL/6 (Jerez-Timaure et al., 2005), M16 \times ICR (Allen et al. 2005), Cast/Ei \times C57BL/6 $-/-$ hg (Farber and Medrano, 2007; Mehrabian et al., 1998), NMRI8 \times DBA/2 (Brockmann et al., 2004; Neuschl et al., 2007), and 129P3 \times C57BL/6 (Reed et al., 2003, 2006) overlap with the QTLs analyzed in our study. Additional QTLs were identified on chromosome 2 outside the QTLs analyzed in this work in the above-mentioned crosses. The dense SNP information of the parental mouse lines of those crosses could help to further evaluate the candidate genes. Fine mapping of QTLs by analysis of animals that are recombinant in the target genomic region will also help to dissect QTL regions physically. The list of genes that we suggest here may also serve as candidate genes for association mapping between SNPs and phenotypes in human populations.

CONCLUSIONS

Data integration has been recognized as one of the key topics since the advent of highthroughput methods in biology. The datasets gained in such highthroughput studies tend to be very comprehensive, if not complete, however little specific. In other words, the jewels, if there are any in the data, are hidden among lots of irrelevant data, and it is hard to discriminate between the two. Here, we have shown that data integration is to some extent possible even for datasets that were gained independently. We could match QTL data and SNP data for 22 murine QTLs for fat weight and body fat percentage from eight independent studies (out of a total of 117 QTLs from 34 published studies) (Wuschke et al., 2006). This allowed us to filter out genes lying in those parts of a QTL where genotypes differed. About half of the nearly 6000 genes in the 22 QTLs could be rejected due to this criterion. Whereas this step of the analysis can be considered direct data integration because different data were combined for one and the same subjects, namely the mouse lines, the other filtering steps that we applied are less evident because we integrated data and knowledge gained for different species. It is very difficult to assess the implications of the homology between a mouse gene and a fatness-related worm gene. However, such indirect data integration has to be applied to reduce further the large number of genes still left after the first filtering step. Introducing scores for features that are an indicator, but not absolutely necessary, for a gene to be associated with fatness, is a simple means to gather further evidence for each gene. We gave the approximately 3000 genes in informative parts of QTLs scores if they showed homology to fatness-related worm genes, had a GO annotation, that is indicating a role in energy metabolism, were highly expressed in liver, hypothalamus, brain, or adipocytes, or if they were shared by more than one QTL. Each of these criteria would rapidly reduce the number of genes if we would require the genes to fulfil strictly all criteria. We found only one gene, *Ganc*, fulfilling all four criteria. We thus conclude that for indirect data integration somewhat softer criteria have to be applied. We realized, furthermore, that the resulting list of genes depends to some extent on the chosen thresholds that are applied for the criteria (e.g., the E-value for a BLAST-hit) and on the scoring scheme. More

elaborate scoring schemes considering the different degrees of importance of the criteria would certainly improve the quality of the candidate gene judgement and selection process. It also became clear that no single evaluation scheme for the analysis of genes within QTLs exist, but a large portion of the analysis is very specific. In the case of this study, the specific parts were the homology searches against worm genes, the expression analysis in a set of selected tissues that are critical for fatness accumulation, and the occurrence of a specific set of keywords in the GO annotation terms. In contrast, the requirement for genes to lie in a QTL region where genotypes differ and the presence in more than one QTL would be the more general criteria that could also be applied to QTL analysis for other traits.

ACKNOWLEDGMENTS

We are grateful to Jonathan Flint and Richard Mott, as well as to the Wellcome Trust, for the generation of the SNP data sets. We thank the Institute of the Novartis Research Foundation, San Diego (USA) for providing the expression data and especially Hilmar Lapp and John Walker for valuable support. Furthermore, we appreciate technical assistance from Axel Rasche and Ralf Herwig, Max-Planck-Institute for Molecular Genetics, Berlin. G.A.B. and H.A.H. were supported by the German Ministry of Education and Research through the National Genome Research Network (NGFN, Grants 01GS0486 and 01GS0487).

SUPPLEMENTARY MATERIAL

Further 121 candidate genes are listed which have two scoring points out of four (GO-annotation, homology, expression, and presence in multiple QTLs; for details, see Materials and Methods). All genes lie in regions of fatness related murine QTLs where genotypes differ.

REFERENCES

- ALLAN, M.F., EISEN, E.J., and POMP, D. (2005). Genomic mapping of direct and correlated responses to long-term selection for rapid growth rate in mice. *Genetics* **170**, 1863–1877.
- ALTSCHUL, S.F., MADDEN, T.L., SCHAFER, A.A., ZHANG, J., ZHANG, Z., MILLER, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- ASHBURNER, M., BALL, C.A., BLAKE, J.A., BOTSTEIN, D., BUTLER, H., CHERRY, J.M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29.
- ASHRAFI, K., CHANG, F.Y., WATTS, J.L., FRASER, A.G., KAMATH, R.S., AHRINGER, J., et al. (2003). Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**, 268–272.
- BROCKMANN, G.A., and BEVOVA, M.R. (2002). Using mouse models to dissect the genetics of obesity. *Trends Genet.* **18**, 367–376.
- BROCKMANN, G.A., KARATAYLI, E., HALEY, C.S., RENNE, U., ROTTMANN, O.J., and KARLE, S. (2004). QTLs for pre- and postweaning body weight and body composition in selected mice. *Mamm Genome* **15**, 593–609.
- BROWN, A.C., OLVER, W.I., DONNELLY, C.J., MAY, M.E., NAGGERT, J.K., SHAFFER, D.J., et al. (2005). Searching QTL by gene expression: analysis of diabetes. *BMC Genet* **6**, 12, PMID: 15760467.
- CERVINO, A.C., LI, G., EDWARDS, S., ZHU, J., LAURIE, C., TOKIWA, G., et al. (2005). Integrating QTL and high-density SNP analyses in mice to identify *Insig2* as a susceptibility gene for plasma cholesterol levels. *Genomics* **86**, 505–517.
- DARVASI, A., and SOLLER, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199–1207.
- DiPETRILLO, K., WANG, X., STYLIANOU, I.M., and PAIGEN, B. (2005). Bioinformatics toolbox for narrowing rodent quantitative trait loci. *Trends Genet* **21**, 683–692.
- DURAND, E., BOUTIN, P., MEYRE, D., CHARLES, M.A., CLEMENT, K., DINA, C., et al. (2004). Polymorphisms in the amino acid transporter solute carrier family 6 (neurotransmitter transporter) member 14 gene contribute to polygenic obesity in French Caucasians. *Diabetes* **53**, 2483–2486.
- EPSTEIN, L.H., JARONI, J.L., PALUCH, R.A., LEDDY, J.J., VAHUE, H.E., HAWK, L., et al. (2002). Dopamine transporter genotype as a risk factor for obesity in African-American smokers. *Obes Res* **10**, 1232–1240.

- FARBER, C.R., and MEDRANO, J.F. (2007) Fine mapping reveals sex bias in quantitative trait loci affecting growth, skeletal size and obesity-related traits on mouse chromosomes 2 and 11. *Genetics*, **175**, 349–360.
- FLINT, J., VALDAR, W., SHIFMAN, S., and MOTT, R. (2005). Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet* **6**, 271–286.
- GAO, Z., WANG, Z., ZHANG, X., BUTLER, A.A., ZUBERI, A., GAWRONSKA-KOZAK, B., et al. (2007). Inactivation of PKC leads to increased susceptibility to obesity and dietary insulin resistance in mice. *Am J Physiol Endocrinol Metab* **292**, E84–E91.
- GAUTIER, L., COPE, L., BOLSTAD, B.M., and IRIZARRY, R.A. (2004). affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315.
- HITZEMANN, R., MALMANGER, B., REED, C., LAWLER, M., HITZEMANN, B., COULOMBE, S., et al. (2003). A strategy for the integration of QTL, gene expression, and sequence analyses. *Mamm Genome* **14**, 733–747.
- HORVAT, S., BÜNGER, L., FALCONER, V.M., MACKAY, P., LAW, A., BULFIELD, G., and KEIGHTLEY, P.D. (2000). Mapping of obesity QTLs in a cross between mouse lines divergently selected on fat content. *Mamm Genome* **11**, 2–7.
- ISHIMORI, N., LI, R., KELMENSEN, P.M., KORSTANJE, R., WALSH, K.A., CHURCHILL, G.A., FORSMAN-SEMB, K., and PAIGEN, B. (2004). Quantitative trait loci that determine plasma lipids and obesity in C57BL/6J and 129S1/SvImJ inbred mice. *J Lipid Res* **45**, 1624–1632.
- JEREZ-TIMAURE, N.C., EISEN, E.J., and POMP, D. (2005). Fine mapping of a QTL region with large effects on growth and fatness on mouse chromosome 2. *Physiol Genomics* **21**, 411–422.
- KASPRZYK, A., KEEFE, D., SMEDLEY, D., LONDON, D., SPOONER, W., MELSOPP, C., et al. (2004). EnsMart: a generic system for fast and flexible access to biological data. *Genome Res* **14**, 160–169.
- KLEIN, T.W. (1978). Analysis of major gene effects using recombinant inbred strains and related congenic lines. *Behav Genet* **8**, 261–268.
- LAN, H., CHEN, M., FLOWERS, J., YANDELL, B., STAPLETON, D., MATA, C., et al. (2006). Combined expression trait correlations and expression quantitative trait locus mapping. *PLoS Genet* **2**, e6.
- LE BACQUER, O., PETROULAKIS, E., PAGLIALUNGA, S., POULIN, F., RICHARD, D., CIANFLONE, K., et al. (2007). Elevated sensitivity to diet-induced obesity and insulin resistance in mice lacking 4E-BP1 and 4E-BP2. *J Clin Invest* **117**, 387–396.
- LEMBERTAS, A.V., PERUSSE, L., CHAGNON, Y.C., FISLER, J.S., WARDEN, C.H., PURCELL-HUYNH, D.A., DIONNE, F.T., GAGNON, J., NADEAU, A., LUSIS, A.J., and BOUCHARD, C. (1997). Identification of an obesity quantitative trait locus on mouse chromosome 2 and evidence of linkage to body fat and insulin on the human homologous region 20q. *J Clin Invest* **100**, 1240–1247.
- MEHRABIAN, M., WEN, P.Z., FISLER, J., DAVIS, R.C., and LUSIS, A.J. (1998). Genetic loci controlling body fat, lipoprotein metabolism, and insulin levels in a multifactorial mouse model. *J Clin Invest* **101**, 2485–2496.
- MEHRABIAN, M., ALLAYEE, H., STOCKTON, J., LUM, P.Y., DRAKE, T.A., CASTELLANI, L.W., et al. (2005). Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nat Genet* **37**, 1224–1233.
- NAGAO, M., PARIMOO, B., and TANAKA, K. (1993). Developmental, nutritional, and hormonal regulation of tissue-specific expression of the genes encoding various acyl-CoA dehydrogenases and alpha-subunit of electron transfer flavoprotein in rat. *J Biol Chem* **268**, 24114–24124.
- NEUSCHL, C., BROCKMANN G.A., and KNOTT S.A. (2007). Multiple trait QTL mapping for body and organ weights in a cross between NMRI8 and DBA/2 mice. *Genet Res* **89**, 47–59.
- PEI, Z., JIA, Z., and WATKINS, P.A. (2006). The second member of the human and murine bubblegum family is a testis- and brainstem-specific acyl-CoA synthetase. *J Biol Chem* **281**, 6632–6641.
- PERUSSE, L., RANKINEN, T., ZUBERI, A., CHAGNON, Y.C., WEISNAGEL, S.J., ARGYROPOULOS, G., et al. (2005). The human obesity gene map: the 2004 update. *Obes Res* **13**, 381–490.
- PLETCHER, M.T., McCLURG, P., BATALOV, S., SU, A.L., BARNES, S.W., LAGLER, E., et al. (2004). Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol* **2**, e393.
- R DEVELOPMENT CORE TEAM. (2005). *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria).
- REED, D.R., LI, X., MCDANIEL, A.H., LU, K., LI, S., TORDOFF, M.G., et al. (2003). Loci on chromosomes 2, 4, 9, and 16 for body weight, body length, and adiposity identified in a genome scan of an F2 intercross between the 129P3/J and C57BL/6ByJ mouse strains. *Mamm. Genome* **14**, 302–313.
- REED, D.R., MCDANIEL, A.H., LI, X., TORDOFF, M.G., and BACHMANOV, A.A. (2006). Quantitative trait loci for individual adipose weights in C57BL/6ByJ × 129P3/J F2 mice. *Mamm Genome* **17**, 1065–1077.
- REIFSNYDER, P.C., CHURCHILL, G., and LEITER, E.H. (2000). Maternal environment and genotype interact to establish diabetes in mice. *Genome Res* **10**, 1568–1578.

- REN, D., ZHOU, Y., MORRIS, D., LI, M., LI, Z., and RUI, L. (2007). Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. *J Clin Invest* **117**, 397–406.
- RUBI, B., DEL ARCO, A., BARTLEY, C., SATRUSTEGUI, J., and MAECHLER, P. (2004). The malate-aspartate NADH shuttle member Aralar1 determines glucose metabolic fate, mitochondrial activity, and insulin secretion in beta cells. *J Biol Chem* **279**, 55659–55666.
- SAKAI, T., SAKAUE, H., NAKAMURA, T., OKADA, M., MATSUKI, Y., WATANABE, E., et al. (2007). Skp2 controls adipocyte proliferation during the development of obesity. *J Biol Chem* **282**, 2038–2046.
- STAJICH, J.E., BLOCK, D., BOULEZ, K., BRENNER, S.E., CHERVITZ, S.A., DAGDIGIAN, C., et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* **12**, 1611–1618.
- STEINBERG, S.J., MORGENTHAUER, J., HEINZER, A.K., SMITH, K.D., and WATKINS, P.A. (2000). Very long-chain acyl-CoA synthetases. Human “bubblegum” represents a new family of proteins capable of activating very long-chain fatty acids. *J Biol Chem* **275**, 35162–35169.
- STYLIANOU, I.M., CHRISTIANS, J.K., KEIGHTLEY, P.D., BÜNGER, L., CLINTON, M., BULFIELD, G., et al. (2004). Genetic complexity of an obesity QTL (Fob3) revealed by detailed genetic mapping. *Mamm Genome* **15**, 472–481.
- STYLIANOU, I.M., CLINTON, M., KEIGHTLEY, P.D., PRITCHARD, C., TYMOWSKA-LALANNE, Z., BÜNGER, L., et al. (2005). Microarray gene expression analysis of the Fob3b obesity QTL identifies positional candidate gene Sqle and perturbed cholesterol and glycolysis pathways. *Physiol Genomics* **20**, 224–232.
- SU, A.I., COOKE, M.P., CHING, K.A., HAKAK, Y., WALKER, J.R., WILTSHIRE, T., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA* **99**, 4465–4470.
- SUVIOLAHTI, E., OKSANEN, L.J., OHMAN, M., CANTOR, R.M., RIDDERSTRALE, M., TUOMI, T., et al. (2003). The SLC6A14 gene shows evidence of association with obesity. *J Clin Invest* **112**, 1762–1772.
- TAYLOR, B.A. and PHILLIPS, S.J. (1997). Obesity QTLs on mouse chromosomes 2 and 17. *Genomics* **43**, 249–257.
- TAYLOR, B.A., TARANTINO, L.M., and PHILLIPS, S.J. (1999). Gender-influenced obesity QTLs identified in a cross involving the KK type II diabetes-prone mouse strain. *Mamm Genome* **10**, 963–968.
- TAYLOR, B.A., WNEK, C., SCHROEDER, C., and PHILLIPS, S.J. (2001). Multiple obesity QTLs identified in an intercross between the NZO (New Zealand obese) and the SM (small) mouse strains. *Mamm Genome* **12**, 95–103.
- WAGENER, A., SCHMITT, A.O., AKSU, S., SCHLOTE, W., NEUSCHL, C., BROCKMANN, G.A. (2006). Genetic, sex and diet effects on body weight and obesity in unique Berlin Fat Mouse inbred lines. *Physiol Genomics* **27**, 264–270.
- WALDER, K., McMILLAN, J., LAPSYS, N., KRIKETOS, A., TREVASKIS, J., CIVITARESE, A., et al. (2002). Calpain 3 gene expression in skeletal muscle is associated with body fat content and measures of insulin resistance. *Int J Obes Relat Metab Disord* **26**, 442–449.
- WANG, X., KORSTANJE, R., HIGGINS, D., and PAIGEN, B. (2004). Haplotype analysis in multiple crosses to identify a QTL gene. *Genome Res* **14**, 1767–1772.
- WILLIAMS, R.W., GU, J., QI, S., and LU, L. (2001). The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biol* **2**, PMID: 11737945.
- WEST, D.B., GOUDEY-LEFEVRE, J., YORK, B., and TRUETT, G.E. (1994). Dietary obesity linked to genetic loci on chromosomes 9 and 15 in a polygenic Mouse model. *J Clin Invest* **94**, 1410–1416.
- WUSCHKE, S., DAHM, S., SCHMIDT, C., JOOST, H.G., and AL-HASANI, H. (2007). Meta-analysis of quantitative trait loci associated with body weight and adiposity in mice. *Int J Obes (Lond)* **31**, 829–841.
- YAMAUCHI, T., NIO, Y., MAKI, T., KOBAYASHI, M., TAKAZAWA, T., IWABU, M., et al. (2007). Targeted disruption of AdipoR1 and AdipoR2 causes abrogation of adiponectin binding and metabolic actions. *Nat Med* **13**, 332–339.
- YU, T.T., McINTYRE, J.C., BOSE, S.C., HARDIN, D., OWEN, M.C., and McCLINTOCK, T.S. (2005). Differentially expressed transcripts from phenotypically identified olfactory sensory neurons. *J Comp Neurol* **483**, 251–262.

Address reprint requests to:
 Armin O. Schmitt
 Institute for Animal Sciences
 Humboldt-Universität zu Berlin
 Invalidenstraße 42
 10115 Berlin, Germany

E-mail: armin.schmitt@agrar.hu-berlin.de